

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-288547

(43) 公開日 平成9年(1997)11月4日

(51) Int.Cl.<sup>8</sup>

G 0 6 F 3/06

識別記号

5 4 0

3 0 1

庁内整理番号

F I

G 0 6 F 3/06

5 4 0

3 0 1 J

技術表示箇所

審査請求 未請求 請求項の数 4 F D (全 15 頁)

(21) 出願番号 特願平8-124088

(22) 出願日 平成8年(1996)4月22日

(71) 出願人 000153454

株式会社日立インフォメーションテクノロ  
ジー

神奈川県秦野市堀山下1番地

(72) 発明者 吉田 徹

神奈川県秦野市堀山下1番地 株式会社日  
立コンピュータエレクトロニクス内

(72) 発明者 青木 健

神奈川県秦野市堀山下1番地 株式会社日  
立コンピュータエレクトロニクス内

(72) 発明者 館下 昌司

神奈川県秦野市堀山下1番地 株式会社日  
立コンピュータエレクトロニクス内

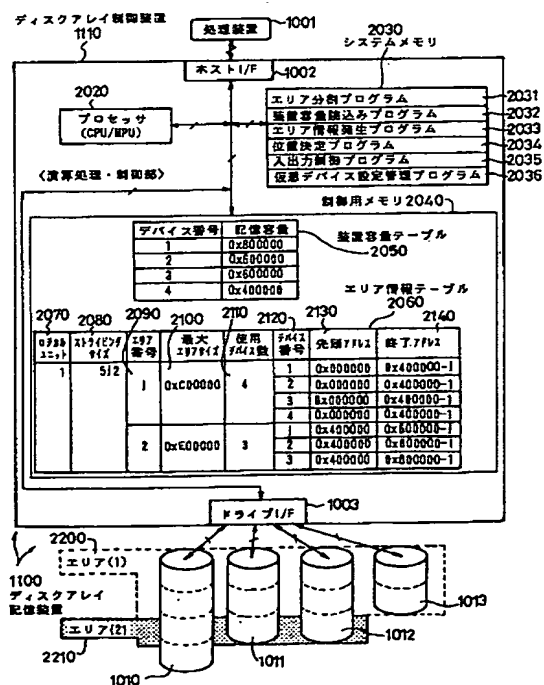
(74) 代理人 弁理士 梶山 信是 (外1名)

(54) 【発明の名称】 アレイ型記憶装置

(57) 【要約】 (修正有)

【課題】複数の記憶デバイスの1つに冗長データを記憶し、複数のデバイスに共通のアドレス空間を割当てて単一の記憶装置として管理するアレイ型記憶装置において、搭載する記憶デバイスの記憶容量の制限をなくすること。

【解決手段】複数の記憶デバイスのそれぞれの最大記憶容量を読み出す手段と、記憶デバイスの記憶容量のうち最も小さい第1の最大記憶容量に合わせて第1のアドレス空間を割当てアレイを形成する第1の記憶エリアを設定し、かつ、最も小さい第1の最大記憶容量の記憶デバイスを除いた複数の記憶デバイスにおいてそれぞれの最大記憶容量から最も小さい第1の最大記憶容量分を差引いた残りのうち最も小さい第2の最大記憶容量に合わせて第2のアドレス空間を割当てアレイを形成する第2の記憶エリアを設定するエリア設定手段と、上位処理装置からのアクセスアドレスに基づいてアレイを形成する第1または第2の記憶エリアに従って記憶デバイスの物理的なアドレスを算出するアドレス算出手段とを備える。



(2)

特開平9-288547

## 【特許請求の範囲】

【請求項1】複数の記憶デバイスの1つに冗長データを記憶し、複数のデバイスに共通のアドレス空間を割当てて単一の記憶装置として管理するアレイ型記憶装置において、前記複数の記憶デバイスのそれぞれの最大記憶容量を讀出す手段と、前記記憶デバイスの記憶容量のうち最も小さい第1の最大記憶容量に合わせて第1の前記アドレス空間を割当てアレイを形成する第1の記憶エリアを設定し、かつ、前記最も小さい第1の最大記憶容量の前記記憶デバイスを除いた複数の前記記憶デバイスにおいてそれぞれの最大記憶容量から前記最も小さい第1の最大記憶容量分を差引いた残りのうち最も小さい第2の最大記憶容量に合わせて第2の前記アドレス空間を割当てアレイを形成する第2の記憶エリアを設定するエリア設定手段と、上位処理装置からのアクセスアドレスに基づいて前記アレイを形成する前記第1または第2の記憶エリアに従って前記記憶デバイスの物理的なアドレスを算出するアドレス算出手段とを備えるアレイ型記憶装置。

【請求項2】前記複数の記憶デバイスのうちの2以上のデバイスを1つの仮想記憶デバイスとして処理する仮想記憶デバイス設定手段を備え、前記エリア設定手段は、この仮想記憶デバイス設定手段により仮想記憶デバイスとされた前記2以上の記憶デバイス以外のデバイスも仮想記憶デバイスとして各前記仮想記憶デバイスについて前記第1および第2のエリア設定をするものであり、前記アドレス算出手段は、前記仮想記憶デバイスに基づいて前記記憶デバイスの物理的なアドレスを算出するものである請求項1記載のアレイ型記憶装置。

【請求項3】複数の記憶デバイスと、上位処理装置のデータ書込み要求に対して前記上位処理装置の書込みデータに冗長データを付加し、前記書込みデータと冗長データを前記記憶デバイスに分散して書込み、前記上位処理装置からのデータ讀出し要求に対して必要なデータを前記記憶デバイスから讀出して前記上位処理装置に転送し、かつ、障害などにより前記記憶デバイス上のデータに直接アクセスできない場合に前記冗長データを用いてアクセスできないデータを復元する制御装置を有するアレイ型記憶装置において、前記記憶デバイスの単体容量を前記制御装置に讀出す讀出手段と、前記讀出手段で讀出した前記記憶デバイスの単体容量に基づいて前記複数の記憶デバイスに所定のアドレス空間を割当て複数のアレイとして管理する管理手段と、前記上位処理装置からアクセスされたときに、前記上位処理装置からのアクセスアドレスに基づいて前記複数のアレイの1つを形成する前記記憶デバイスの物理的なアドレスを算出するアドレス算出手段とを備えるアレイ型記憶装置。

【請求項4】前記管理手段は、前記記憶デバイスに対してアレイとして管理する際に、2台以上の記憶デバイスを仮想的な1台の記憶デバイスとして管理するものであ

り、前記アドレス算出手段は、前記上位処理装置からアクセスされたときに、前記上位処理装置からのアクセスアドレスに基づいて前記仮想的な記憶デバイスの論理的なアドレスを算出する論理アドレス算出手段と前記論理的アドレス算出手段で算出された論理的なアドレスから前記記憶デバイスの物理的なアドレスを算出する物理的アドレス算出手段とを有するアレイ型記憶装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】この発明は、アレイ型記憶装置において、ファイルなどのまとまりのあるデータを複数の磁気ディスク記憶装置等の外部記憶装置（以下記憶デバイスという）に分散して冗長データを付加して記憶するアレイ型の記憶装置に関し、詳しくは、容量の異なる記憶デバイスを複数使用した場合においても、アレイ形式で効率よく管理することができるアレイ型の記憶装置に関する。

## 【0002】

【従来の技術】従来のアレイ形式の記憶装置として磁気ディスク記憶装置を多数配列してこれらを一体的な単一の記憶領域として扱うディスクアレイ・システムが一般的に知られている。ディスクアレイ・システムは、計算機システムにおいて、多数の磁気ディスク装置を並列に入出力動作させることにより、処理装置と記憶デバイスとの間で高速なデータ転送を行い、高い信頼性の実現をするシステムである。

【0003】ディスクアレイ・システムの構成の例としては、Paterson等の論文(D. Paterson, G. Gibson, R. Katz, "A Case for Redundant Arrays of Inexpensive Disks (RAID)", ACM SIGMOD conference proceedings, 1988, pp.109-116)に記載がある。この論文では、RAID (レイド, Redundant Array of Inexpensive Disks) と呼ばれるディスクアレイ形態を提示している。RAIDは、通常の入出力データを複数の記憶デバイスに分散記憶すると共に、冗長データを記憶する。この冗長データの設置は、前記データに欠落が生じた場合に、データ復元処理を行うために用いられ、これにより、欠落したデータを復元することを可能にする。

## 【0004】

【発明が解決しようとする課題】図7は、アレイ型記憶装置の1つとして一般的に知られているディスクアレイ記憶装置を用いるディスクアレイ記憶システムの機能ブロック図である。図7は、アレイ型の記憶システム（ディスクアレイ記憶システム）を示していて、これは、入出力要求を発行する、プロセッサを有する処理装置1001（いわゆるホストコンピュータ）と各種のデータを記憶するディスクアレイ記憶装置1030とからなる。ディスクアレイ記憶装置1030は、ディスクアレイ制御装置1000とデータを記憶する磁気ディスク記憶装

(3)

特開平9-288547

置（以下ディスク記憶デバイスという。）1010～1013等から構成されている。ディスクアレイ制御装置1000は、処理装置1001と物理的に接続されるホストインタフェース（以下インタフェースはI/Fと略す。）1002、制御用メモリ1004、ディスク記憶デバイス1010～1013との間で物理的な接続を行うドライブI/F1003、そして冗長データの生成、データの読出／書込制御等の各種の制御を行なうための演算処理・制御部（CPU＋メモリ）1005とにより構成される。

【0005】前記の構成によりディスクアレイ制御装置1000は、処理装置1001からのアクセスに対して、ディスク記憶デバイス装置1010～1013との入出力処理、冗長データの生成、冗長データを用いたデータの復元処理などを行なう。また、ロジカルユニット1020は、ディスク記憶デバイス1010～1013を処理装置1001からみて論理的に1台の記憶装置デバイスとして認識される記憶領域を示すものであって、1021は、ディスク記憶デバイス1010～1013において処理装置1001から1台の記憶装置として認識されない未使用領域である。

【0006】このようなディスクアレイ記憶装置にあっては、記憶されるデータの冗長性を維持し、かつ搭載される記憶デバイスの記憶容量を最大限に利用するために、搭載する記憶デバイスの各々の記憶容量を同一にする必要がある。そのため、図7に示すように搭載される記憶デバイスの中に一つでも記憶容量の少ない記憶装置があった場合には、記憶容量が最小の記憶デバイスを基準にしてアレイ型記憶装置1030が構成される。それが論理的に構築する1つのディスク記憶デバイス（ロジカルユニット）1020である。すなわち、最小の記憶容量を持つディスク記憶デバイスより多い記憶容量を持つディスク記憶デバイスは、最小記憶容量との差分の記憶容量が未使用（未使用領域1021）となり、この記憶領域（記憶単体）が無駄になる問題がある。

【0007】また、アレイ型記憶装置において、記憶容量増加の目的で増設されるディスク記憶デバイスも、増設したディスク記憶デバイスを最大限に利用するためには、増設以前に搭載されているディスク記憶デバイスの記憶容量を考慮して同一な記憶容量のものを用意することが必要になる。この発明の目的は、このような従来技術の問題点を解決するものであって、データを複数の記憶デバイスに分散して記憶するアレイ型記憶装置において、記憶デバイスの記憶容量を考慮しなくても済み、搭載する記憶デバイスの記憶容量の制限をなくすることができるアレイ型記憶装置を提供することにある。

【0008】

【課題を解決するための手段】このような目的を達成するためのこの発明のアレイ型記憶装置の特徴は、複数の記憶デバイスの1つに冗長データを記憶し、複数のデバ

イスに共通のアドレス空間を割当てて単一の記憶装置として管理するアレイ型記憶装置において、複数の記憶デバイスのそれぞれの最大記憶容量を読出す手段と、記憶デバイスの記憶容量のうち最も小さい第1の最大記憶容量に合わせて第1のアドレス空間を割当てアレイを形成する第1の記憶エリアを設定し、かつ、最も小さい第1の最大記憶容量の記憶デバイスを除いた複数の記憶デバイスにおいてそれぞれの最大記憶容量から最も小さい第1の最大記憶容量分を差引いた残りのうち最も小さい第2の最大記憶容量に合わせて第2のアドレス空間を割当てアレイを形成する第2の記憶エリアを設定するエリア設定手段と、上位処理装置からのアクセスアドレスに基づいてアレイを形成する第1または第2の記憶エリアに従って記憶デバイスの物理的なアドレスを算出するアドレス算出手段とを備えるものである。

【0009】

【発明の実施の形態】前記の構成のように、この発明にあっては、複数の記憶デバイスの単体容量が異なる場合にあっては、上位処理装置から1台の記憶装置として認識されるエリア（記憶単位）を持つ、異なったアレイ複数形成する。さらに、他の発明として、アレイを形成する際には、2台以上の記憶デバイスを仮想的な1台の記憶デバイスと見なして、上位処理装置からのアクセスされた際に、上位処理装置からのアクセスアドレスから、仮想的な記憶デバイスの論理的なアドレスを特定し、特定した論理的なアドレスから前記記憶デバイスの物理的なアドレスを特定するようにする。このようにすることで、制御装置の配下に接続される記憶デバイスの単体容量が異なる場合であっても、複数のアレイ領域を持たせることで上位処理装置から1台の記憶装置として認識されるようにすることができる。その結果、異なる記憶容量を持つ記憶デバイスを使用しても、上位処理装置から1台の記憶装置として認識される容量効率の高いアレイ形式の記憶装置を実現することができる。なお、アレイ型記憶装置では、ストライピングサイズ（後述）で分散記憶をするために、冗長データを記憶するデバイスは、複数の記憶デバイスの1つではあるが、この1つの記憶デバイスは、データの書込みの都度決定され、その都度異なるものになる。

【0010】

【実施例】図1は、この発明のアレイ型記憶装置を適用した一実施例のディスクアレイ記憶システムの機能ブロック図である。図1において、1100は、ディスクアレイ記憶装置であって、ディスクアレイ制御装置1110とディスク記憶デバイス1010～1013とにより構成され、ディスクアレイ制御装置1110は、演算処理・制御部2010と制御メモリ2040等を有している。演算処理・制御部2010は、プロセッサ（CPUあるいはMPU）2020と各種処理プログラムを記憶したシステムメモリ2030からなる。制御メモリ2040

(4)

特開平9-288547

は、図7の制御用メモリ1004に対応する制御用メモリである。システムメモリ2030には、エリア分割プログラム2031と装置容量読み込みプログラム2032、エリア情報発生プログラム2033、位置決定プログラム2034、入出力制御プログラム2035、仮想デバイス設定管理プログラム2036等が設けられている。

【0011】装置容量読み込みプログラム2032は、ディスクアレイ記憶装置を構築する際に、ドライブI/F1003に接続された各ディスク記憶デバイス1010～1013の記憶容量を読み込み、制御用メモリ2040にディスク記憶デバイス番号別に記憶容量（その記憶デバイスの最大記憶容量）のデータをテーブルの形式（装置容量テーブル2050）で格納する処理プログラムである。このプログラムは、その終了時点で次のエリア分割プログラム2031をコールする。なお、このとき、ディスク記憶デバイス1010～1013は、デバイス番号1～4が割当てられ、この番号で管理される。エリア分割プログラム2031は、前記のディスク記憶デバイス番号別に作成された記憶容量データを元にして複数のデバイス間においてアドレス空間が共通するエリアを単位として順次エリア分割して複数のエリアを論理的に発生させて番号付けをし、アレイとして発生した各エリアを管理する処理プログラムである。ここでは、ディスク記憶デバイス1010～1013の記憶容量を元にして2つのエリア(1)2200とエリア(2)2210とを発生させ、各エリアに対応させてディスク記憶デバイス情報をテーブル2060の形式で制御用メモリ2040に形成する。

【0012】エリア情報発生プログラム2033は、処理装置1001からのアクセス要求ブロックアドレス（ストライピングサイズ、例えば512バイトごとに分割されて領域をアクセスするので、各アドレスはブロックアドレスになる。）を元にして前記のテーブル2060を参照してアクセスすべき分割されたエリアを特定するエリア情報（エリア(1)か、エリア(2)か）を発生する処理プログラムである。なお、処理装置1001は、アクセス要求ブロックアドレスのほかに、記憶情報についてのサイズ情報も送出する。このサイズ情報に応じてエリアが次にまたがる場合には、そのエリア情報も発生する。また、前記のストライピングサイズは、データを各デバイスごとに分散記憶する場合のデータ記憶の単位であって、これはデータ転送の単位として管理される領域、いわゆる1セクタのバイト数に相当するものである。位置決定プログラム2034は、処理装置1001からのアクセス要求ブロックアドレス及び前記エリア情報を元にして、計算式により冗長データを格納するディスク記憶デバイス、データを格納するディスク記憶デバイス及びアクセス要求データのディスク記憶デバイスにおけるブロックアドレス（物理アドレス）を後述する計

算式によりそれぞれ算出する処理プログラムである。記憶容量テーブル2050は、装置容量読み込みプログラム2032により形成され、制御用メモリ2040上に作成されるデバイスごとの、それぞれの最大記憶容量を記憶するテーブルである。エリア情報テーブル2060は、エリア分割プログラム2031により形成され、制御用メモリ2040上に作成されるエリアの詳細情報を記憶するテーブルである。

【0013】エリア情報テーブル2060の欄2070は、ロジカルユニット番号の格納場所である。欄2080は、ディスクアレイ装置構築の際に指定するストライピングサイズについて格納する場所である。欄2090は、分割されたエリアの番号を格納する場所である。欄2100は、一つのエリアで使用されている論理空間（アドレス空間）の最大エリアサイズ格納場所である。欄2110は、一つのエリア内で使用されているディスク記憶デバイスの個数を格納する場所である。欄2120はデバイス番号、欄2130は先頭物理アドレス、欄2140は最終物理アドレスをそれぞれディスク記憶デバイスに対応させて記憶する欄である。2200は、エリア分割プログラム2031により生成されたディスク記憶デバイス1010～1013を対象として分割して論理的に形成された第1のアレイエリアを構成する1つ目のエリア(1)であり、2210は、ディスク記憶デバイス1010～1013を対象として分割して論理的に形成された第2のアレイエリアを構成する2つ目のエリア(2)である。

【0014】まず、装置容量読み込みプログラム2032は、各ディスク記憶デバイス1010～1013に対してRead Capacityコマンドを発行して、ディスク記憶デバイス毎の記憶容量を読み込み、制御用メモリ2040内に装置容量テーブル2050を設けてデバイス番号に対応させてそれぞれの記憶容量を記憶し、その処理を終了する。エリア分割プログラム2031は、装置容量読み込みプログラムの終了により、コールされて実行され、次に説明するエリア分割処理とエリア情報テーブル2060の生成処理を行う。

【0015】前記の処理について図2のフローチャートに従って説明すると、まず、エリア分割プログラム2031は、始めにロジカルユニット番号をエリアテーブル2060内のロジカルユニット番号格納場所である欄2070に格納し（ステップ3010）、これから構築しようとするディスクアレイ装置のストライピングサイズをストライピングサイズ格納場所2080に格納した後（ステップ3020）、ディスク記憶デバイス単体で使用済みの記憶容量格納変数 $z$ を初期化（ステップ3030）し、エリア情報テーブルの内容を作成するために装置情報テーブル2050をメモリ2030の作業領域にコピーする処理を行う（ステップ3040）。

【0016】次に、最小記憶容量のディスク記憶デバイ

(5)

特開平9-288547

ス検索するためにデバイス番号1(1010)の記憶容量を変数 $x$ に格納する(ステップ3050)。変数 $x$ に格納された値と次のディスク記憶デバイスの記憶容量とを比較して(ステップ3060)、比較した結果、次のディスク記憶デバイスの記憶容量の方が小さい場合は次のディスク記憶デバイスの記憶容量を変数 $x$ に格納する(ステップ3070)。変数 $x$ に格納されていた値の方が多い場合は、ステップ3060で比較した次のディスク記憶デバイスの、その次にも比較すべきディスク記憶デバイスが存在するかを判定する(ステップ3080)。これは、コピーした装置容量テーブルにステップ3060で比較した次のディスク記憶デバイスの、その次のディスク記憶デバイス記憶容量が“0”になっていないかをチェックすることによる。もしも、記憶容量が“0”になっている場合には、つまり次に比較すべきディスク記憶デバイス存在しない場合には、エリア番号1をエリア番号格納場所2090に格納する(ステップ3090)。

【0017】次に、エリア(1)での構成ディスク記憶デバイス数を求めるために、変数 $y$ にエリア(1)の構成ディスク記憶デバイス数の初期値1を格納する(ステップ3100)。コピーした装置容量テーブルで記憶容量が“0”でない装置全てを対象とし、前記の最小記憶容量検索処理で求められた最小容量の値 $x$ 分を減算した記憶容量をコピーした装置容量テーブルの各ディスク記憶デバイス対応に書込む(ステップ3110)。言い換えれば、ここで、エリア分割済み記憶容量を削減する処理をする。そこで、コピーした装置容量テーブルには、記憶容量“0”のディスク記憶デバイスと残りの記憶容量が“0”でないディスク記憶デバイスが登録されている。このステップ3110で削減した状態において、次のディスク記憶デバイスの記憶容量が“0”になっていないかをチェックする(ステップ3130)。次のディスク記憶デバイスがある場合はエリア(1)の構成ディスク記憶デバイス数が格納されている変数 $y$ を1カウントアップする(ステップ3120)。そして、次にディスク記憶デバイスがなくなるまでこれを続けて、ディスク記憶デバイスがなくなった場合、変数 $y$ の値をエリアテーブル2060内のディスク記憶デバイス数格納場所2110に格納する(ステップ3140)。

【0018】次に、エリア(1)で使用している各ディスク記憶デバイスの使用空間情報、つまり、各ディスク記憶デバイスの先頭ブロックアドレス(物理アドレス)と最終ブロックアドレス(物理アドレス)をエリア情報テーブル2060において、それぞれのディスク記憶デバイスのデバイス番号に対応させて欄2130、欄2140に格納する処理を行なう。まず、先頭アドレスは合計使用記憶容量に対応して決定される値 $z$ を格納する(ステップ3150)が、このときの使用容量は“0”になっているので、“0”が記録されることになる。エリア

(1)は、最初に構築されるものであり、また、合計使用記憶容量 $z$ は初期化されているので、どのディスク記憶デバイスも先頭ブロックアドレスは“0”である。

【0019】各ディスク記憶デバイスの使用最終アドレスは、各ディスク記憶デバイスの先頭ブロックアドレスに最小記憶容量検索処理で求められた最小容量値を加算した値、つまり、(先頭ブロックアドレス値 $=z$ ) + 変数 $x$ となるので、その値を格納する(ステップ3160)。変数 $y$ に格納されているエリア(1)での使用ディスク記憶デバイス数分、すなわち、使用ディスク記憶デバイス数 $y$ から1を減算した値が0以上(ステップ3170)である間は、前記のステップ3150からステップ3170をループして、エリアテーブルの装置別使用空間格納場所2140~2146に順次先頭アドレスと最終アドレスとを格納していく(ステップ3180)。

【0020】次に、合計使用記憶容量 $z$ に $z$  + 変数 $x$ を代入して、これを次の先頭アドレス $z$ として記憶する(ステップ3180)。ここまでの処理でエリア(1)を構築するためのエリア情報テーブル2060を作成することができる。次に、コピーした装置容量テーブルのディスク記憶デバイス記憶容量が“0”になっているディスク記憶デバイスの数が1か否かを判定し(ステップ3190)、2以上であるときには、ステップ3050へと戻り、次のエリア分割処理を同様な処理を繰り返すことで行う。これをディスク記憶デバイスの数が1になるまでくり返し行ない、繰り返しの都度エリア番号をインクリメントすることでエリア情報テーブル2060が完成し、エリア分割が終了する。これによりエリア(1)とエリア(2)とによる2つのアレイ記憶領域が形成される。なお、ディスク記憶デバイス記憶容量が“0”になっているディスク記憶デバイスの数が1であるときには、冗長データが記憶できないので、このエリアは使用されない。

【0021】次に、複数の記憶装置を物理的に並列接続させた状態の記憶装置を一つの仮想記憶装置として論理的に認識し、アレイ型記憶装置として動作させる仮想記憶装置方式について説明する。これは、記憶容量の小さい記憶デバイスが複数存在するような場合のエリア設定について大きなアドレス空間のエリアを設定することができる点で有効な処理になる。まず、図3を用いて仮想記憶装置について説明する。7000は、物理的に並列接続されたディスク記憶デバイス装置を表わしている。7010は、新規に構築される1個の論理ディスク記憶デバイス(ロジカルユニット)を表わしている。7020は、図7に対応する増設後のディスクアレイ装置の未使用領域を表わしている。7030は、論理的に直列接続された仮想記憶デバイスを表わしている。

【0022】通常アレイ型記憶装置において記憶容量増加の目的で増設する記憶デバイスは、図3の上を示すように、アレイ型記憶装置7000は、実ディスク記憶デ

(6)

特開平9-288547

バイス（以下実デバイスという。）1010～1013（RD1～RD4）に対して新たに増設した実デバイス4080（RD5）とがドライブI/F1003に物理的に並列接続される。しかし、この状態では、未使用領域は、増設前と変わらず無駄な領域ができてしまう。そこで、図3の下に示すように物理的に接続された状態から増設分実デバイス4080（RD5）と実デバイス1013（RD4）とを論理的に直列接続し、さらに論理的に直列接続された実デバイス1013（RD4）と実デバイス4080（RD5）を1つの仮想記憶デバイスとして認識し、論理的に連続した一つの記憶領域（1個のディスク記憶デバイス）として扱う。このデバイスが論理ディスク記憶デバイスとしての仮想デバイス7030である。なお、以上のような仮想デバイスの設定と管理は、仮想デバイス設定管理プログラム2036により行われる。

【0023】図4は、エリア分割方式とこの仮想記憶装置方式を併用した時に、新たに制御用メモリ2040内に作成される仮想記憶装置管理テーブル4000と、装置容量テーブル2050、そしてエリア情報テーブル4010とを示している。これらテーブルは、ディスク記憶デバイスの分割状態及び仮想ディスク記憶デバイス（以下仮想デバイスという。）として認識されている状態を表わしている。仮想記憶装置管理テーブル4000は、実デバイス（RD1～RD5）の記憶容量情報が格納されている装置容量テーブル2050を元に仮想デバイス設定管理プログラム2036により作成される。これには仮想デバイスの詳細情報が格納される。エリア情報テーブル4010は、図2のエリア情報テーブル2060のデバイス数の欄2110が仮想デバイス個数の欄2150になり、デバイス番号欄が仮想デバイス番号（VD）欄2160になり、さらに、RD（実デバイス）の個数欄2170、RD番号欄5180が加えられ、仮想デバイス（VD）と実デバイス（RD）とを管理するテーブルである。

【0024】ここで、エリア4020は、仮想デバイスを含み、エリア分割されたエリア(1)の領域を示す。エリア4030は、仮想デバイス（VD）を含み、エリア分割されたエリア(2)の領域を示す。4040～4070は、それぞれ仮想デバイス（VD）である。また、4080は、ディスクアレイ記憶装置の記憶容量を増加する目的で新たに増設された実デバイス（RD）である。まず始めに、装置容量読み込みプログラム2032により、接続されている実デバイス（RD）の記憶容量を全て読出して、装置容量テーブル2050を制御用メモリ2030に作成する。次に装置容量テーブル2050と、オペレータによって指定される仮想デバイスの個数、この実施例では、仮想デバイスVD1、VD2、VD3、VD4の4個と、オペレータによって指定されるそれぞれの仮想デバイスに割当てた実デバイスの情報、

この実施例では仮想デバイスVD4が実デバイス1013、4080（RD4、RD5）から構成され、そのほかは、それぞれの実デバイスから構成される。これらの情報を入力して、この入力データに応じて仮想デバイス設定管理プログラム2036により仮想記憶装置を管理するエリア情報テーブル4010が制御用メモリ2040内に作成される。次に、仮想デバイス設定管理プログラム2036によりエリア分割プログラム2031がコールされて、1仮想デバイス（VD）を1ディスク記憶デバイスとして扱い、前記実施例のエリア情報テーブル作成手順に従って、1つのエリア内における仮想デバイス単体の記憶容量と、仮想デバイス数と、1つのエリア内の1つの仮想デバイスにおける実デバイスの使用領域情報を求めてエリア情報テーブル4010を制御用メモリ2040内に作成する。このときには、仮想デバイス数の欄2150、仮想デバイス番号（VD）欄2160、RD（実デバイス）の個数欄2170、RD番号欄5180は、それぞれ入力情報に応じてエリア分割プログラム2031により生成され、それぞれの情報が記録される。

【0025】次に、こうして構築したエリア分割方式と仮想記憶装置方式を併用した場合のアレイ型の記憶デバイスを例として上位処理装置1001からのアクセス要求に対する書き込み処理及び読出し処理の動作について図5に従って説明する。図5は、上位処理装置1001からの書き込みアクセス要求に対して実デバイス番号と確定した実デバイス内に格納されているブロックアドレス（物理アドレス）を決定する処理動作を示したものである。5000は、入出力制御プログラム2035による処理装置1001からのアクセス要求受信を示す。その内容の一例として書き込み要求を示している。5010は、処理装置1001から見た、この実施例におけるディスクアレイ装置の論理アドレス空間を示すものである。

【0026】位置決定プログラム2034は、処理装置1001からのアクセス要求を受入れて、指定されたブロックアドレスに対して所定の演算式に従って演算処理をして記憶デバイスを決定してその物理的な位置（アドレス）を算出する。入出力制御プログラム2035は、その位置においてデータの読み／書きを行うとともに、処理装置1001のデータ書き込み要求に対しては処理装置1001の書き込みデータに対して冗長データを生成する処理と、位置決定プログラム2034により決定された位置を基準として、書き込みデータと冗長データとをディスク記憶デバイスに分散して書き込む処理、処理装置1001からのデータ読出し要求に対しては必要なデータを位置決定プログラム2034により決定された位置を基準としてディスク記憶デバイスから読出して、処理装置1001に転送する処理、障害などによりディスクデバイス上のデータに直接アクセスできないような異常に

(7)

特開平9-288547

おいては、冗長データを用いてアクセスできないデータを復元する処理を行うが、冗長データの生成や冗長データによるデータの復元の処理は、アレイ型記憶装置において、従来から行われている処理であるので、その詳細な説明をここでは省略する。また、ディスク記憶デバイスに分散して書込む処理とは、データをストライピングサイズごとに記憶デバイスを更新しながら順次書込む処理である。したがって、所定のサイズのデータは、ストライピングサイズで各記憶デバイスごとに分散して配置される。そして、その次の記憶デバイスに上長データが書込まれる。なお、冗長データは、データ記憶デバイスとは異なる、データの書込みの都度決定される記憶デバイスに記憶されるものである。これを記憶する記憶デバイスがデータを記憶するデバイスのほかに必要になる。そこで、アレイ型記憶装置のデータ記憶容量は、実際にエリア管理されるディスク記憶デバイスよりも1つ少ないデバイスがデータ記憶デバイスになる。また、ここでは、冗長データのシンドロームビットをパリティビットとしてこれを記憶するデバイスをパリティ記憶デバイス（パリティDV）とし、データ記憶デバイスをデータDVとして説明する。

【0027】位置決定プログラム2034は、次の計算式5240に従ってアクセス要求ブロックアドレスのパリティデータが存在する仮想デバイスの番号PNを算出する。

$$PN = VD - ((LBA/S)/D) \% VD$$

ただし、Sは、ストライピングサイズ（符号5200）、VDは、エリア単位の構成仮想デバイス（VD）の数（符号5210）、Dは、データを記憶するエリア単位の構成データVD数（符号5220）であって、 $D = VD - 1$ による。この $D = VD - 1$ により冗長データを除く、実データの記録デバイス数が算出される。LBAは、論理ブロックアドレス値（符号5230）である。また、%は除算した結果の余りを算出する演算子である。LBA/Sにより、論理アドレスについて物理記録単位で換算したカウント総数が算出され、 $(LBA/S)/D$ により実際にデータを書込み/読出しする、あるエリア単位でのデバイスの物理アドレス数が算出される。 $((LBA/S)/D) \% VD$ により実際にデータを書込み/読出する最後のデバイス番号が算出される。その結果として、アクセス要求ブロックアドレスのパリティデータが存在する仮想デバイスが何番目であるかが、前記式により番号PNとして算出される。

【0028】また、位置決定プログラム2034は、次の計算式5250に従ってデータを書込み/読出するデバイス番号（データVD）の情報DDを算出する。

$$DD = (LBA/S) \% D + 1$$

$(LBA/S) \% D$ によりあるエリア単位においてデータを書込み/読出するデバイス最後のデバイス番号が決定され、アクセス要求ブロックアドレスのデータが存在す

る仮想デバイスが何番目であるかが、前記式により情報DDとして算出される。さらに、位置決定プログラム2034は、次の条件付き計算式5260に従ってデータを書込み/読出するデバイス（データVD）のデバイス番号DNを算出する。

$$DN = \text{if } (DD \geq PN)$$

$$DN = DD + 1$$

$$\text{else}$$

$$DN = DD$$

これにより、パリティデータが存在する仮想デバイスを含めて仮想デバイスの番号を算出する。そして、位置決定プログラム2034は、次の計算式5270に従ってアクセスブロックアドレスABAを算出する。

$$ABA = ((LBA/S)/D) \times S + (LBA \% (LBA/S))$$

この式により実デバイスの物理アドレスが算出される。

【0029】次に、処理装置1001から所定のブロックアドレスへの書込みを例として全体的な動作を説明する。入出力制御プログラム2035は、ブロックアドレス $0 \times 1000000$ 番地に書込みアクセス要求を処理装置1001から受信（5000）した場合、エリア情報発生プログラム2033をコールし、このプログラムにより、始めに、エリアテーブルの最大エリアサイズと要求を受けたブロックアドレスからエリア番号を算出する。このプログラムは、エリア(1)の最大エリアサイズ $0 \times 1203100$ 番地（5130）よりアクセス要求ブロックアドレス $0 \times 1000000$ 番地の方が小さいことから、要求ブロックアドレスはエリア(1)内（5120）に存在すると特定する。入出力制御プログラム2035は、次に位置決定プログラム2034をコールして、このプログラムによりストライピングサイズ：S（5200）にはエリア情報テーブル4010内のストライピングサイズ512バイト（5110）を代入する。そして、エリア単位の構成仮想デバイス数：VD（5210）には、エリア(1)の構成仮想デバイス数4（5140）を代入し、エリア単位の構成データ仮想デバイス数：D（5220）には、 $VD - 1 = 3$ を代入する。さらに、アクセス要求ブロックアドレスLBA（5230）には、アクセス要求ブロックアドレスの $0 \times 1000000$ 番地を代入する。

【0030】これらの値をアクセス要求ブロックアドレスが存在するデータのパリティ仮想デバイス番号を特定するための計算式（5240）とアクセス要求ブロックアドレスが存在するデータのディスク記憶デバイスが何番目かを特定する計算式5250に代入してそれぞれの値を算出する。前記計算式により算出されたデータディスク記憶デバイス値：DDは、パリティ仮想デバイス位置によって仮想デバイス位置が変わるため、計算式（5260）によりデータディスク記憶デバイス番号を特定する。前記の実施例のディスクアレイ装置では、PN＝

(8)

特開平9-288547

2, D=3, DN=4となる。

【0031】次に図6に従って、入出力制御プログラム2035による、前記の算出された情報を元にしてアクセスを行なうまでの手順を説明する。始めに、仮想デバイス内の何番目の実デバイスであるかを記憶する変数RDcountを初期する(ステップ6000)。次に、エリアは(1)と判明していることと、前記で算出した情報DNが4であることから、エリア情報テーブル4010のエリア(1)、仮想デバイス番号4、RDcount:1の先頭アドレスを変数xに(ステップ6010)入れ、最終アドレスを変数yに格納する(ステップ6020)。次に、アクセスアドレスを計算式(5270)で算出する。前記の実施例では、0x555400番地という値が算出される(ステップ6030)。算出されたアクセスブロックアドレスABAが最大エリアサイズを越えている可能性があるので、一つ前のエリアの最大エリアサイズを減算する(ステップ6040)。

【0032】エリア(1)へのアクセスであるため“0”をここでは減算する。次に上記で算出したアクセスブロックアドレス(0x555400番地)に先頭ブロックアドレスxを加算した値が最終ブロックアドレスyを越えているか、つまり、アクセスブロックアドレスの値が使用している実デバイス容量を越えていないかチェックする(ステップ6050)。もし越えている場合は、実デバイス先頭アドレス:xにアクセスブロックアドレス:ABAを加算した値からその実デバイスの最終アドレス:yを減算した値をABAに格納し(ステップ6060)、RDcountを1加算する(ステップ6070)処理をアクセス領域を発見するまでくり返す。越えていない場合、つまりアクセス領域を発見した場合は、エリア情報テーブル4010とRDcountから実デバイス番号を求め、変数RDNoに格納する(ステップ6080)。実デバイスへのアクセスは、求められた実デバイスの先頭アドレスxに求められたアクセスブロックアドレスABAを加算した値のアドレスに行なう。

【0033】このようにして得た情報から前記の実施例の場合は、実デバイス番号(RDNo:5)のアクセスブロックアドレス(ABA:0x155400番地)にアクセスすることで処理装置1001からの要求に正しく応答することが可能である(ステップ6090)。なお、処理装置1001からの読み込みアクセス要求に対しても同様の処理を行なうことにより正しいアドレスへのアクセス処理が可能である。以上、この発明を実施例に基づき具体的に説明したが、この発明は、前記実施例に限定されるものではなく、その要旨を逸脱しない範囲で種々変更しうることは言うまでもない。

【0034】この実施例によれば、このように、アレイ型記憶装置において、制御装置の配下に接続された各記憶デバイスの記憶容量を読み出し、異なる記憶容量の記憶デバイスにおいても、エリア分割を用いてエリア分割情

報作成し、さらには、エリア分割方式と記憶デバイスを論理的に直列に接続する仮想記憶デバイスとの併用により、エリア分割情報内に仮想記憶デバイス情報を盛り込み、処理装置からのアクセス要求に対しては、計算式によって実記憶デバイス番号、アクセスアドレスの算出を容易に行なえるので、アレイ型記憶装置において、異なる記憶容量を持つ記憶デバイスを有効に活用することができる。

【0035】

【発明の効果】この発明にあっては、アレイ形式の記憶装置システムにおいて、搭載する記憶デバイス単体の記憶容量が同一であることを考慮してシステムを構築するという制限をなくし、異なる記憶容量を持つ記憶デバイスを最大限に利用することが可能になる。また、記憶容量増加の目的で増設した増設記憶デバイスの記憶容量が増設以前に搭載されている記憶デバイスの記憶容量と同一でなければ最大限、記憶容量を有効に利用できない制限をなくし、増設する記憶デバイスの記憶容量が増設以前に搭載されている記憶デバイスの記憶容量と異なっても記憶デバイスを最大限に活用できる。

【図面の簡単な説明】

【図1】図1は、この発明のアレイ型記憶装置を適用した一実施例のディスクアレイ記憶システムの機能ブロック図である。

【図2】図2は、図1の実施例におけるエリア情報テーブルの生成処理のフローチャートである。

【図3】図3は、図1の実施例において、物理的に並列接続された記憶デバイスを論理的に直列接続した時の仮想記憶デバイスの認識状態の説明図である。

【図4】図4は、異なる記憶容量を持つ記憶装置を論理的に直列に接続した場合の図1の実施例における、新たに制御用メモリ内に作成するエリア情報テーブルとディスク記憶デバイスの分割状態及び仮想デバイスとして認識されている状態の説明図である。

【図5】図5は、上位処理装置からの書き込みアクセス要求に対して実記憶デバイス番号と確定した実記憶デバイス内に格納されているブロックアドレスを決定する説明図である。

【図6】図6は、図1の実施例において、計算式により算出された情報を元に実記憶デバイスに対してのアクセス処理を行なうまでの処理のフローチャートである。

【図7】図7は、一般的なディスクアレイ型記憶システムの全体的な構成図である。

【符号の説明】

1000…ディスクアレイ制御装置、1001…処理装置、1002ホストI/F、1003…ドライブI/F、1004、2030…制御用メモリ、1005…プロセッサ(CPU/MPU)、1010～1013…ディスク記憶デバイス、1020…ロジカルユニット、1021…未使用領域、1030…ディスクアレイ装置、



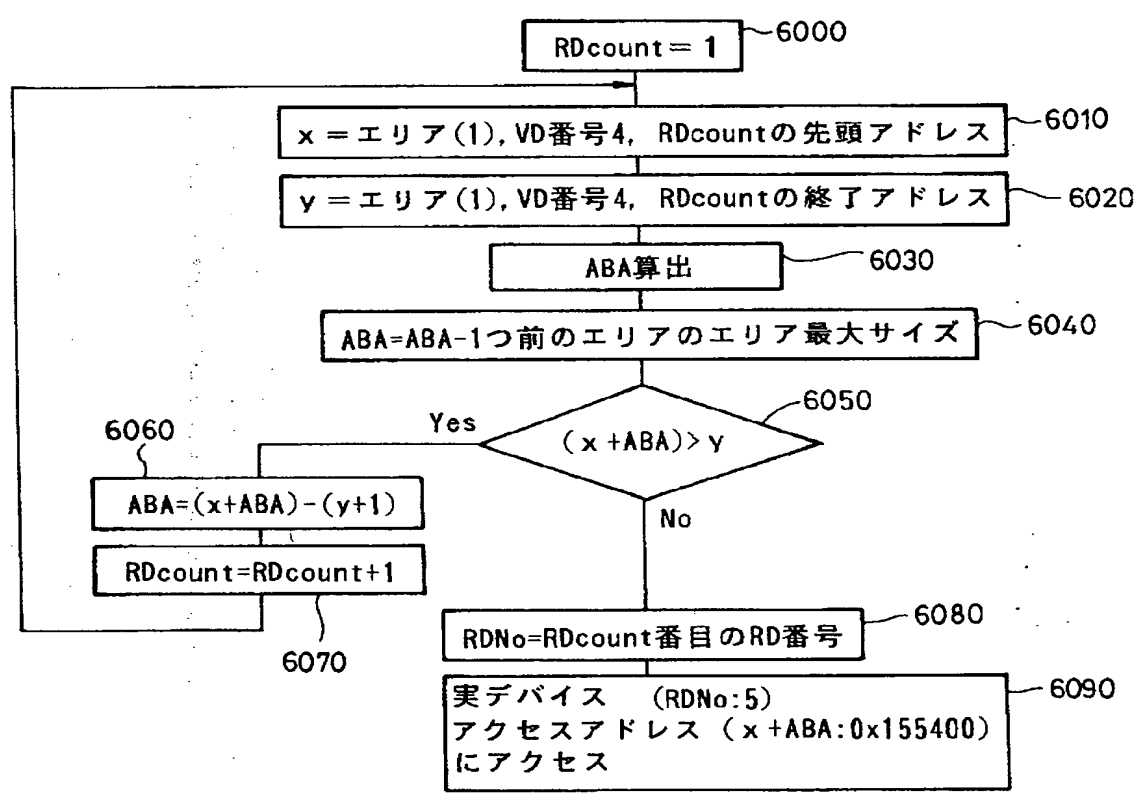
(9)

特開平9-288547

2031…エリア分割プログラム、2031…装置容量読み込みプログラム、2033…エリア識別プログラム、2034…位置決定プログラム、2050…装置容量テーブル、2060…エリア分割情報テーブル、2070…ロジカルユニット番号格納場所、2080…ストライピングサイズ格納場所、2090…エリア番号格納場所、2100…最大エリアサイズ格納場所、2110…使用ディスク記憶デバイス数格納場所、2120～2180…ディスク記憶デバイス使用領域格納場所、2200, 2210…分割したエリア、4000…仮想記憶装

置管理テーブル、4010…仮想記憶装置を考慮したエリア情報テーブル、4020, 4030…仮想記憶装置を考慮し分割したエリア、4040～4070…仮想デバイス、4080…増設したディスク記憶デバイス、5010…処理装置から見た論理空間、5240…パリティディスク記憶デバイス番号算出計算式、5250…データディスク記憶デバイス番号算出計算式、5260…データディスク記憶デバイス番号算出計算式、5270…アクセスブロックアドレス算出計算式、7030…仮想記憶装置。

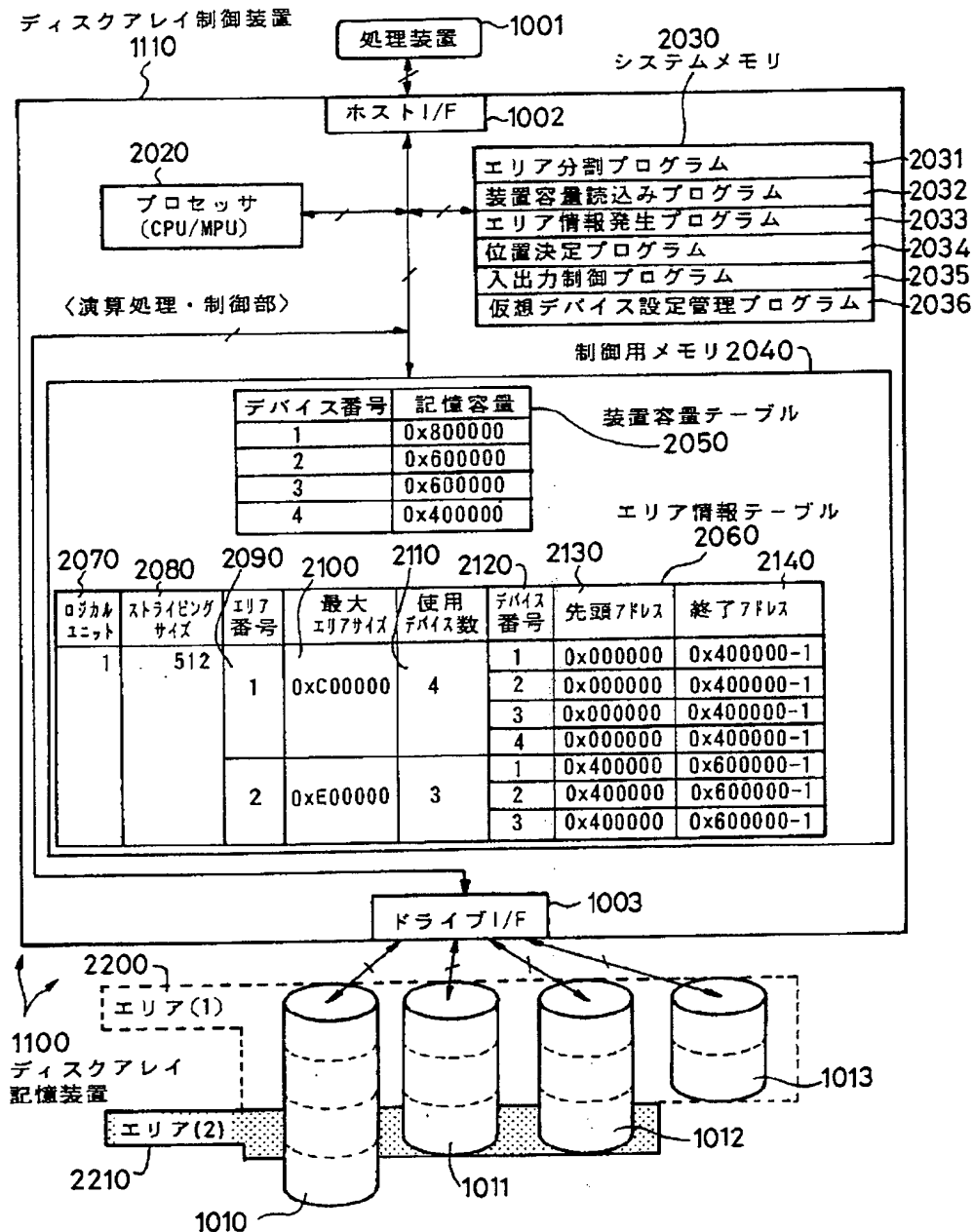
【図6】



(10)

特開平9-288547

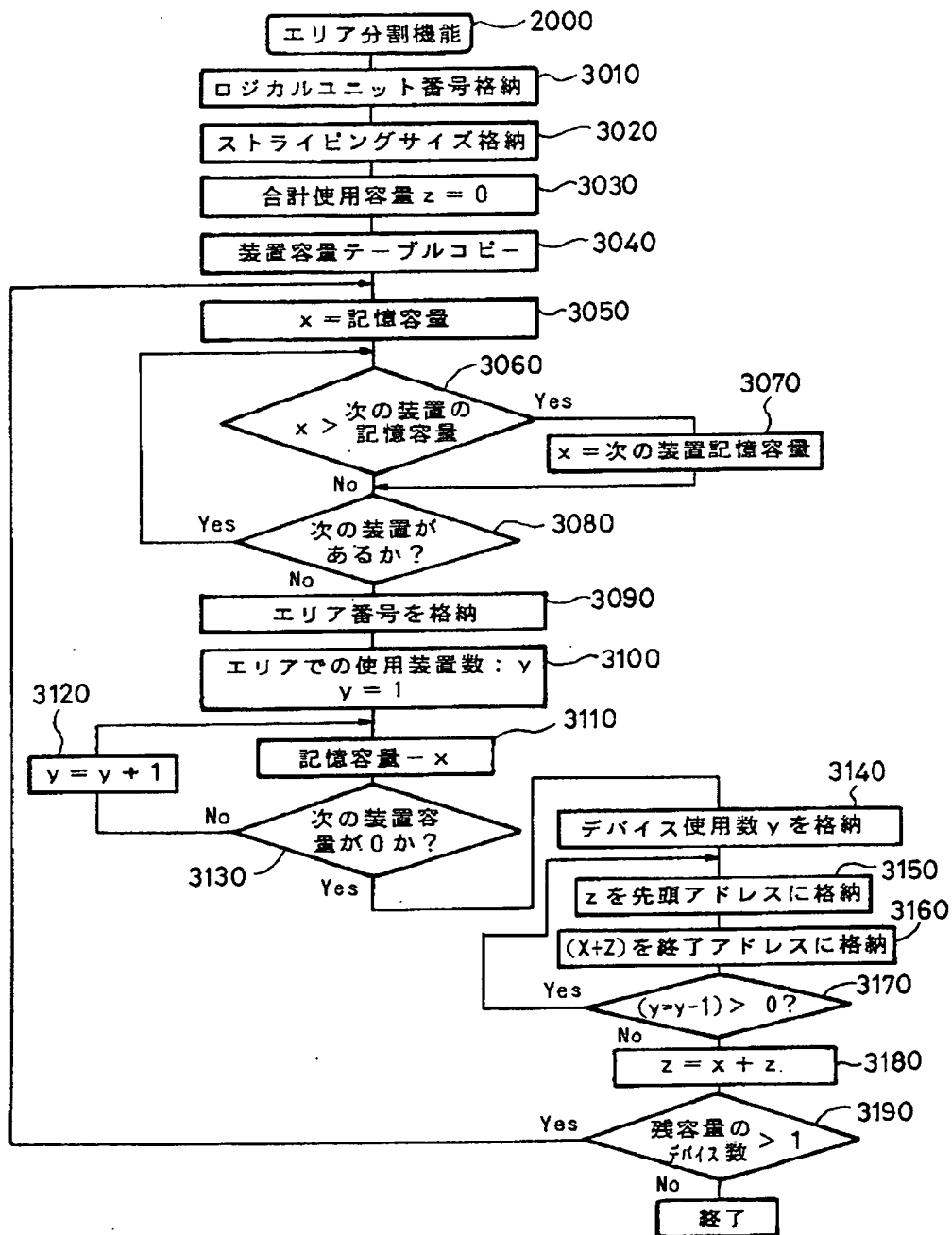
【図1】



(11)

特開平9-288547

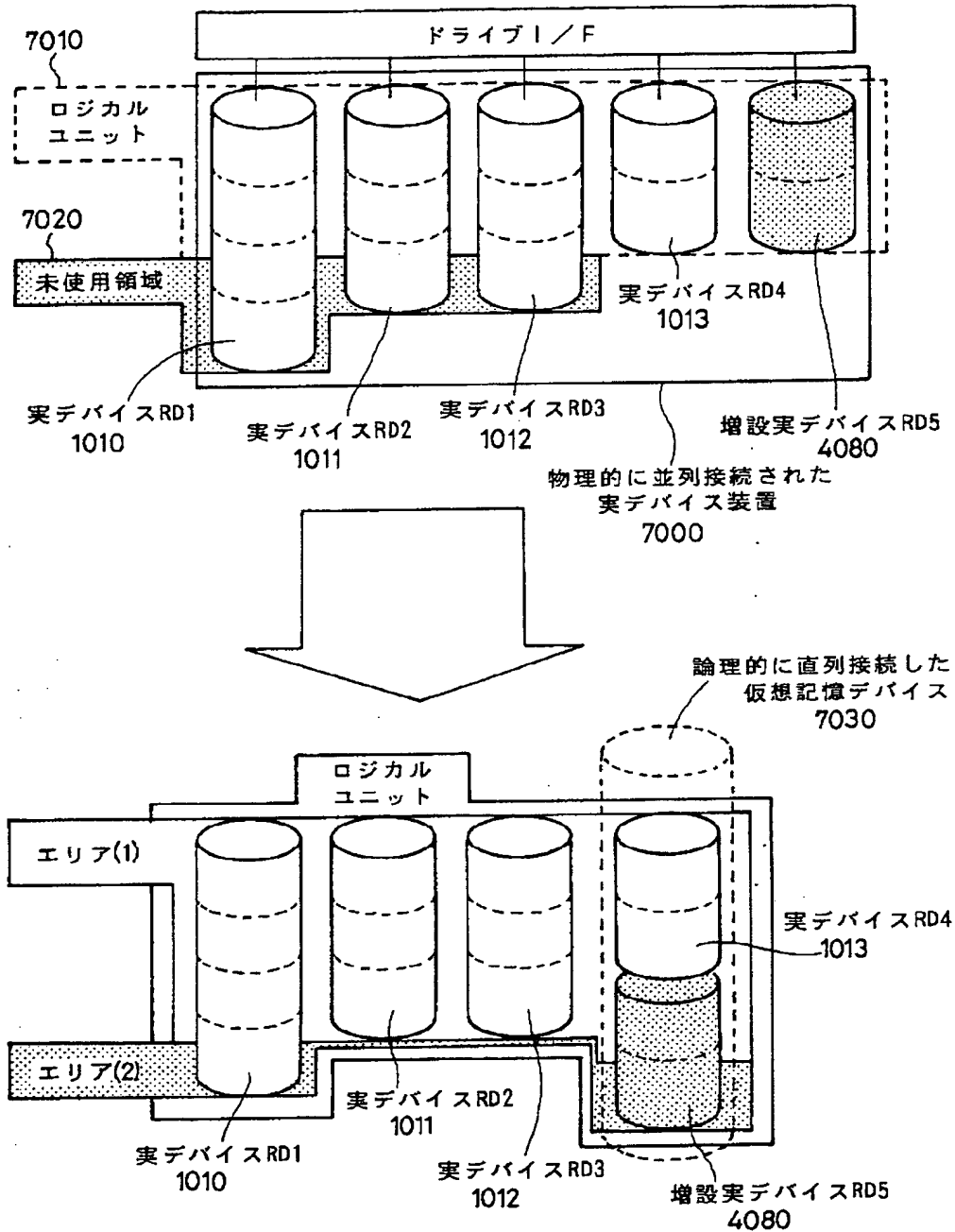
【図2】



(12)

特開平9-288547

【図3】



( 1 3 )

特開平9-288547

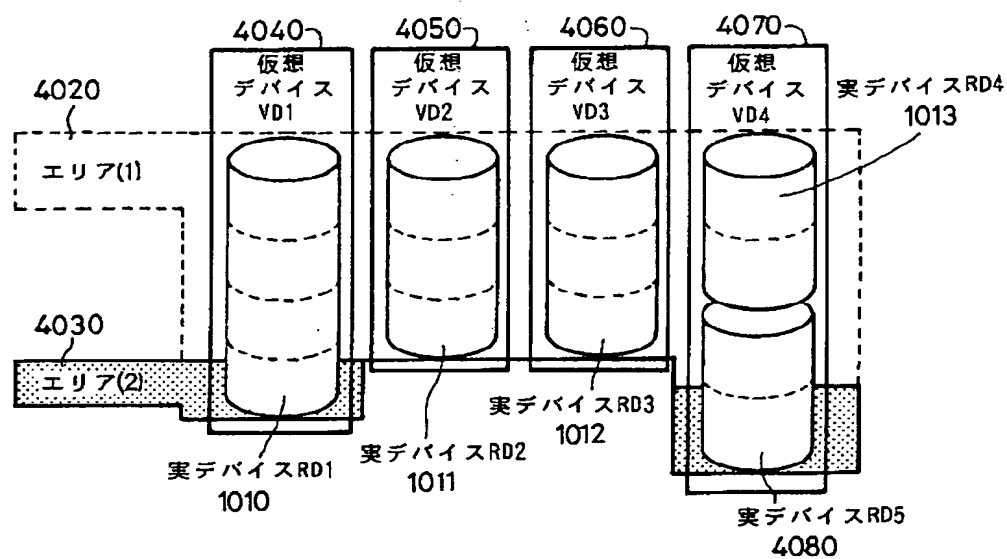
【図4】

制御用メモリ 2040

仮想記憶装置管理テーブル 4000					装置容量テーブル 2050	
VD番号	VD容量	RD個数	RD番号	容量	デバイス番号	記憶容量
1	0x800000	1	1	0x800000	1	0x800000
2	0x600000	1	2	0x600000	2	0x600000
3	0x600000	1	3	0x600000	3	0x600000
4	0x800000	2	4	0x400000	4	0x400000
			5	0x400000		

エリア情報テーブル

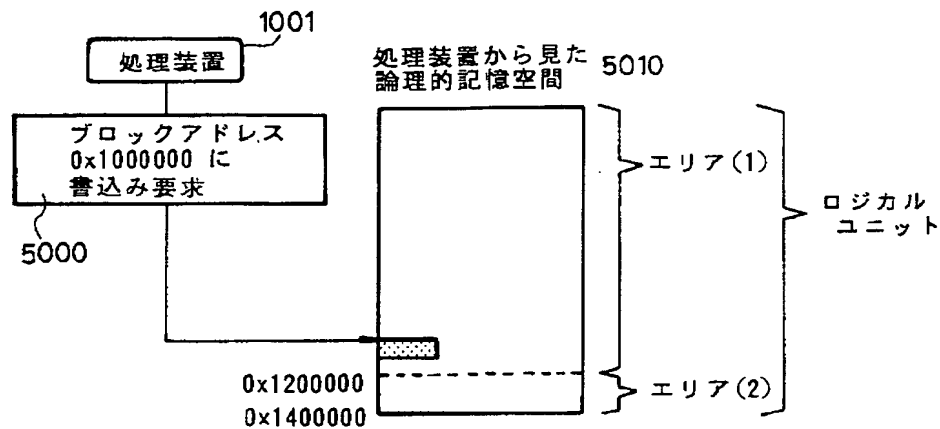
ロジカル ユニット	ストライピング サイズ	エリア 番号	最大 エリアサイズ	仮想デバイス(VD)情報					
				実デバイス(RD)情報					
				VD 個数	VD 番号	RD 個数	RD 番号	先頭アドレス	終了アドレス
1	512	1	0x1200000	4	1	1	1	0x000000	0x600000-1
					2	1	2	0x000000	0x600000-1
					3	1	3	0x000000	0x600000-1
					4	2	4	0x000000	0x400000-1
		2	0x1400000	2			5	0x000000	0x200000-1
					1	1	1	0x600000	0x800000-1
					4	1	5	0x200000	0x400000-1



(14)

特開平9-288547

【図5】



ストライピングサイズ : S ————— 5200  
 エリア単位の構成VD数 : VD ————— 5210  
 エリア単位の構成データVD数: D=VD-1 ————— 5220  
 論理ブロックアドレス : LBA ————— 5230  
 パリティVD番号 :  $PN = VD - (((LBA/S)/D) \% VD)$  ————— 5240  
 データVD :  $DD = (LBA/S) \% D + 1$  ————— 5250  
 データVD番号 :  $DN = \text{if}(DD > PN)$  ————— 5260  
                    $DN = DD + 1$   
                   else  
                    $DN = DD$   
 アクセスブロックアドレス :  $ABA = (((LBA/S)/D) \times S) + (LBA \% (LBA/S))$  ————— 5270

注) A/B: Aを Bで除算した商  
 A%B: Aを Bで除算した余り

エリア情報テーブル 4010

ロジカル ユニット	ストライピング サイズ	エリア 番号	最大 エリアサイズ	仮想デバイス(VD)情報					
				VD 個数	VD 番号	実デバイス(RD)情報			
						RD 個数	RD 番号	先頭アドレス	終了アドレス
1	512	1	0x1200000	4	1	1	1	0x000000	0x600000-1
					2	1	2	0x000000	0x600000-1
					3	1	3	0x000000	0x600000-1
					4	2	4	0x000000	0x400000-1
		2	0x1400000	2		5	5	0x000000	0x200000-1
					1	1	1	0x600000	0x800000-1
					4	1	5	0x200000	0x400000-1

5100

5110

5120

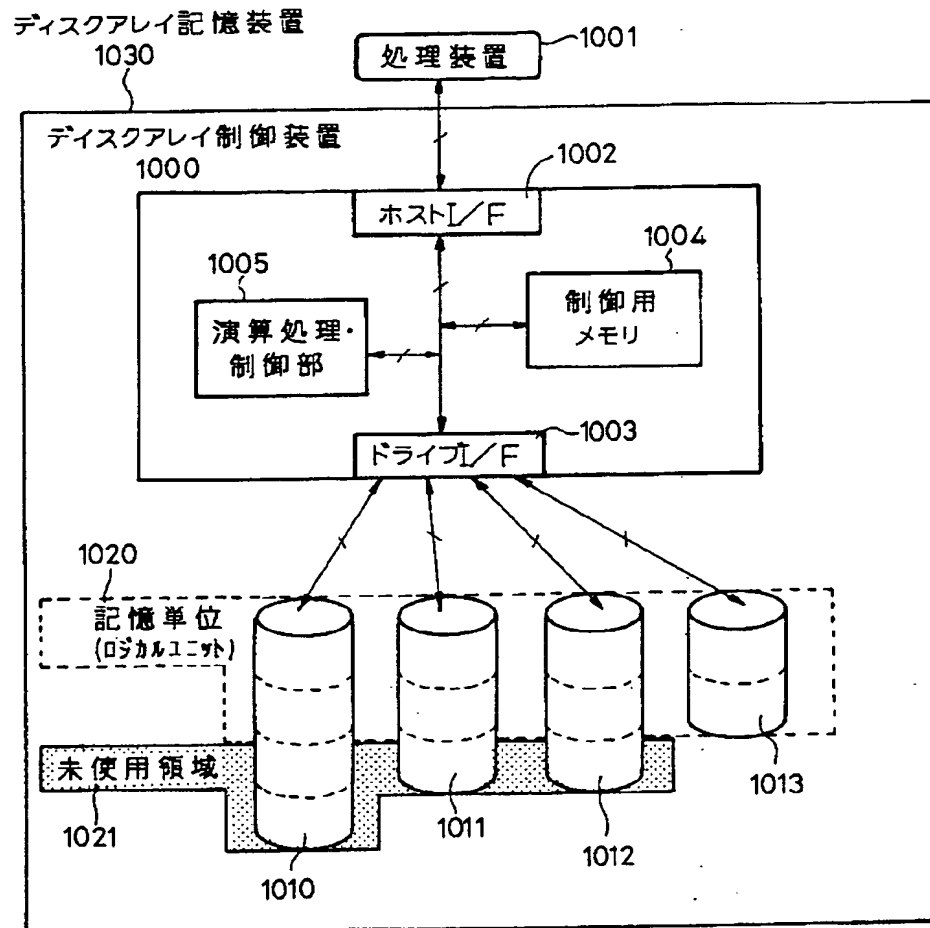
5130

5140

(15)

特開平9-288547

【図7】



**THIS PAGE BLANK (USPTO)**



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**